

*Simón, Lucía; Fontáns, Exequiel; Duarte, Amanda;
Barité, Mario*

Estructuras sintagmáticas en tesauros: aproximación cualitativa-cuantitativa

III Jornadas de Intercambio y Reflexión acerca de la Investigación en Bibliotecología

28 y 29 de noviembre de 2013

CITA SUGERIDA:

*Simón, L.; Fontáns, E.; Duarte, A.; Barité, M. (2013) Estructuras sintagmáticas en tesauros: aproximación cualitativa-cuantitativa [en línea]. III Jornadas de Intercambio y Reflexión acerca de la Investigación en Bibliotecología, 28 y 29 de noviembre de 2013, La Plata, Argentina. En Memoria Académica. Disponible en:
http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.3370/ev.3370.pdf*

Documento disponible para su consulta y descarga en **Memoria Académica**, repositorio institucional de la **Facultad de Humanidades y Ciencias de la Educación (FaHCE)** de la **Universidad Nacional de La Plata**. Gestionado por **Bibhuma**, biblioteca de la FaHCE.

Para más información consulte los sitios:

<http://www.memoria.fahce.unlp.edu.ar>

<http://www.bibhuma.fahce.unlp.edu.ar>



Esta obra está bajo licencia 2.5 de Creative Commons Argentina.
Atribución-No comercial-Sin obras derivadas 2.5

Estructuras sintagmáticas en tesauros: aproximación cualitativa-cuantitativa

Lucía Simón¹, Exequiel Fontás¹, Amanda Duarte¹, Mario Barité¹

¹Universidad de la República, Montevideo, Uruguay. e-mail: lulusimon86@gmail.com

Resumen. De los sistemas organizadores del conocimiento, los tesauros se reconocen por ser los contruidos bajo pautas más refinadas y elaboradas. Los descriptores y no descriptores de un tesoro se dividen, considerando su estructura morfosintáctica, en dos tipos: términos simples (una sola palabra) y términos compuestos (sintagmas terminológicos). Estas formas compuestas pueden ser definidas como términos que tienen una estructura compuesta por dos o más palabras, y representan preferentemente a un solo concepto. La sintagmación es uno de los recursos más comúnmente utilizados para la creación de nuevas unidades terminológicas. Se aventuran algunas explicaciones de esta recurrencia. Por su naturaleza, los sintagmas terminológicos plantean diversos problemas metodológicos: ofrecen dificultades para su reconocimiento formal y su delimitación dentro de un texto especializado; tienden a presentar variaciones lexicales, en especial cuando todavía no están suficientemente consolidadas en una lengua; y, obligan al desarrollo de herramientas complementarias para facilitar la indización o la recuperación temática. Investigaciones precedentes en áreas diferentes del saber determinan que entre el 50 y el 70 % de los términos de un dominio especializado adoptan formas sintagmáticas en la lengua española. Se propone una clasificación de las estructuras formales adoptadas por las unidades terminológicas en cinco categorías, y se plantea un estudio de los sintagmas conforme a esa categorización y a su extensión. Se analizan dos tesauros de las Ciencias Sociales en las dos dimensiones mencionadas. Se analizan los resultados, se establecen conclusiones y se sugieren, a partir de los mismos, tres líneas de investigación complementarias en la materia.

1. Tesauros y control de vocabulario

Los tesauros son sistemas totalmente estructurados, integran con términos que guardan entre sí relaciones semánticas y funcionales, y su propósito principal es proporcionar representaciones lingüísticas para la indización de documentos y recursos de información en general, de modo de facilitar la recuperación de la información, especialmente en áreas especializadas. Los tesauros son totalmente estructurados pues todos y cada uno de los términos tienen un lugar propio dentro de la estructura conceptual y relacional. Los términos no están aislados ya que tienen vínculos con al menos uno de los otros, y el sistema conceptual “refleja la ordenación y la estructura de algún campo del conocimiento.” (Naumis Peña, 2007, p. 129).

De los sistemas organizadores del conocimiento, los tesauros se reconocen por ser los contruidos bajo pautas más refinadas y elaboradas (Gil Leiva, 1999; Dodebei, 2002; National Information Standards Organization, 2005; Mazzocchi *et al*, 2007), pues

cumplen en forma rigurosa con los procesos de control de vocabulario, en tres sentidos diferentes: a) control formal, para establecer formas normalizadas de los términos elegidos; b) discriminación entre términos descriptores, que se autorizan como términos de indización, y términos no descriptores, no autorizados, que solo tienen valor de referencia; y, c) sistematización y control de las relaciones existentes entre los términos. Para lograr el control de vocabulario “as linguagens documentárias incorporam procedimentos de normalização gramatical e semântica”. Mencionan que “a normalização gramatical refere-se à forma de apresentação dos seus elementos quanto ao gênero (geralmente masculino), ao número (uso de singular ou plural) e ao grau” en tanto que “a normalização semântica procura garantir a univocidade na representação dos conceitos de áreas de especialidade” (Cintra et al, 2002, p. 43). Mientras la normalización gramatical apunta a formalizar la expresión externa y visible de un concepto (un término, un símbolo codificado), la normalización semántica procura establecer una significación propia y precisa para cada término.

Los descriptores y no descriptores de un tesoro se dividen, considerando su estructura morfosintáctica, en dos tipos: a) términos simples, que se constituyen con una sola y única palabra, y b) términos compuestos, que se integran con dos o más palabras y que aparecen frecuentemente en la literatura especializada como estructuras fijas y estables (Cabré, 1993; Corbalán Sánchez y Amat, 2003; Krieger & Finatto, 2004; Naumis Peña, 2007). Las expresiones compuestas que en los textos de especialidad no son meros sintagmas de discurso, constituyen “formaciones lexemáticas con valor terminológico” (Cabré, 1993, p. 186), y adquieren valor de referencia por dos motivos: en primer lugar, porque pasan a representar a un concepto dentro de un ámbito del saber; y en segundo lugar, porque ese valor de representación las vuelve aptas para su integración a diccionarios, glosarios y tesauros especializados. Toda expresión compuesta con valor terminológico y de referencia es, entonces, un sintagma terminológico, que aspira a constituirse en un término de indización para tesauros.

La sintagmación es un “recurso formal de obtención de unidades léxicas” que “se basa en la formación de una nueva unidad a partir de una combinación sintáctica jerarquizada de palabras” (Cabré, 1993, p. 189). El diccionario patrón de la lengua española define a los sintagmas como “grupos”, esto es, como conjuntos “de palabras estructuradas, relacionadas en torno a un núcleo”, a las que habitualmente les

corresponde “un comportamiento sintáctico unitario” (Real Academia Española, 2001, p. 1.161 y 2.071). Para Dubuc (1999, p. 56) son “un conjunto de palabras vinculadas por una relación sintáctica identificable”. Cuentan con: a) “una vertiente sistemática (formal, semántica e funcional), toda vez que son unidades de un código establecido”; y, b) una vertiente pragmática, por ser “unidades usadas en la comunicación especializada para designar los objetos de una realidad preexistente”. (Cabré, 1993, p. 169).

Las expresiones sintagmáticas terminológicas comparten con las unidades fraseológicas de la lengua general o de los ámbitos especializados, algunas de sus características más relevantes; siguiendo a Corpas (1996): a) son estables en gran medida y tienden a fijarse de una determinada manera, más allá de presentarse en el discurso con variantes lexicales o sintácticas menores; b) tienen una frecuencia alta de aparición conjunta, lo que distingue a estas expresiones de asociaciones casuales de palabras dentro de un texto; c) tienden a institucionalizarse, como consecuencia del uso frecuente, la fijación y su estabilidad en el tiempo. Los sintagmas terminológicos destinados a la indización cuentan con elementos diferenciales respecto a las llamadas unidades fraseológicas especializadas (Bevilacqua, 2001; L'Homme, 2001), elementos que no serán tratados en este trabajo.

Los sintagmas terminológicos son muy frecuentes en los discursos especializados, constituyéndose, como fue dicho, a partir de la lexicalización de estructuras sintácticas. Son unidades léxicas complejas que cargan un todo de sentido e identifican a un concepto en el ámbito de un lenguaje de especialidad. De acuerdo con Cabré (1993, p. 177), “se rigen por las mismas reglas combinatorias de los sintagmas libres, y, aparentemente, no se distinguen por ningún trazo formal específico”. De allí que no resulte sencillo distinguir entre unos y otros. Algunos sintagmas terminológicos dan un sentido particular a dos o más palabras que separadas no significan lo mismo (“nicho ecológico”). Otros desambiguan expresiones vagas (“ruido documental”) o especifican denominaciones genéricas para aumentar la precisión y la comprensión (“desprendimiento de retina”). Lo que importa es que todo sintagma terminológico se crea por necesidad lingüística y de comunicación en los ámbitos especializados.

La sintagmación es uno de los recursos más comúnmente utilizados para la creación de nuevas unidades terminológicas (Alves, 1998; Krieger y Finatto, 2004). Suele utilizarse para bautizar nuevos objetos, hechos o fenómenos a partir de palabras ya existentes,

antes que asumir el albur de crear nuevas palabras. Opera como un mecanismo por el cual el núcleo del compuesto determina y hace visible la categoría a la que el concepto pertenece, en tanto que el determinante indica el criterio de subdivisión de esa categoría y ofrece un trazo distintivo que carga de significado al nuevo concepto (Sager et al., 1980, p. 265).

Por su naturaleza, los sintagmas terminológicos plantean diversos problemas metodológicos: a) presentan dificultades para su reconocimiento formal y su delimitación dentro de un texto especializado. Cabré, Estopá y Lorente (1996, p. 3) opinan que

en la difícil frontera entre un término de estructura sintagmática y una combinación con un alto valor de aparición, solo los criterios cuantitativos, los formales, aducidos por el análisis lingüístico, y el reconocimiento por la competencia permitirán establecer diferenciaciones con un cierto rigor y precisión

b) tienden a presentarse con variaciones lexicales, en especial cuando todavía no están suficientemente consolidadas en una lengua, o cuando se realizan varias traducciones de un texto relevante a una misma lengua de destino.

En lo que se refiere a los procesos de formación de términos, Sager (1993) diferencia entre medios analíticos y sintéticos de formación; los primeros combinan unidades léxicas independientes para formar unidades mayores (como, por ejemplo, en la composición), en tanto que los segundos modifican los elementos léxicos mediante el uso de afijos. De esa manera, las unidades terminológicas complejas serían el resultado de un proceso analítico de formación de términos. Dicho autor afirma que el método sistemático de designación se basa en la consistencia de la creación de términos mediante la selección de calificadores que indiquen alguna propiedad o otras características esenciales de un concepto. Así, “los modelos regulares de los términos compuestos en teoría podrían desarrollarse con uno, dos, tres y más elementos para indicar las dependencias jerárquicas” (Sager, 1993, p. 101).

En el área de Organización del Conocimiento, la presencia de sintagmas terminológicos presenta ventajas y dificultades. Entre las primeras, permiten “una mayor especificidad para la representación del conocimiento” y aumentan “la precisión porque la heterogeneidad de la pre-coordinación a veces no soluciona al usuario la ubicación de la información que busca en el documento.” (Naumis Peña, 2007, p. 177). Su desventaja

más importante es que por regla general, cuanto mayor es el número de palabras que compone un descriptor, mayores son las posibilidades de que en entornos electrónicos, se pueda recuperar información con alto nivel de *ruido* (entendido como “conjunto de documentos o informaciones que resultan ser no pertinentes o no relevantes” a una demanda de información, al momento de la recuperación (Barité, 1997, p. 126), o de *fuga de datos* (pérdida o invisibilidad de información pertinente a una demanda y que no es recuperada a partir de las claves temáticas del usuario demandante). Estas dificultades se matizan por la posibilidad de combinar descriptores o palabras sueltas, en una base de datos bien estructurada. De todas maneras, la extensión de un descriptor afecta siempre en alguna medida el acceso y recuperación de un documento (Lancaster, 1996).

El caudal de sintagmas ha provocado el surgimiento o desarrollo de herramientas complementarias de control de vocabulario, autónomas o adosadas a tesauros, para facilitar la indización o la recuperación temática. Una de ellas es el índice de permutación o índice permutado, que ordena alfabéticamente los descriptores y no descriptores de un tesoro, los que aparecen tantas veces como palabras significativas tenga el término. Así, la expresión “tasa de deserción escolar” aparecerá tres veces en el orden alfabético del índice permutado, bajo “deserción”, “escolar” y “tasa”. El índice permutado facilita la ubicación de sintagmas cuando el indizador y/o el usuario conocen o intuyen que puede encontrar variaciones de un término.

Por otra parte, los estándares internacionales relacionados con la creación y la gestión de vocabularios controlados incluyen normas específicas para el tratamiento de sintagmas terminológicos. La norma norteamericana (National Information Standards Organization, 2005) y la británica (British Standards Institution, 2005) dan cuenta de la preocupación por regular y normalizar la creación de términos compuestos, ante la evidencia del alto número de sintagmas terminológicos detectable en cualquier área del conocimiento.

Investigaciones precedentes hacen presumir que entre el 50 y el 70 % de los términos de un dominio especializado adoptan formas sintagmáticas en lengua castellana y en lengua portuguesa. Así, en una pesquisa realizada en el área de Contratación Comercial Internacional, se comprobó que el 64,3 % de los 801 términos identificados, se constituían como sintagmas (Barité et al, 2005). Krieger y Finatto (2004, p. 81)

aseguran que “há vários estudos que comprovam a prevalência das terminologias instituídas ao modo de sintagmas, num porcentual que se situa em torno de 70 % das ocorrências terminológicas”. Barité y Simón (2010) compararon cuatro repertorios terminológicos de los que resultó que el 60,4 % de los términos registrados en ellos configuraban sintagmas. No obstante, pueden encontrarse excepciones, como en el trabajo de Gil Leiva y Rodríguez Muñoz (1997), lo que obliga a continuar estudios contrastivos en esta línea. Tanto la Norma ISO (1987) como Krieger y Finatto (2004) mencionan que el alto porcentaje de sintagmas terminológicos se da sobre todo en áreas nuevas del conocimiento o en aquellas donde es evidente el surgimiento de neologismos.

En este trabajo se propone una aproximación cuantitativa y cualitativa a la incidencia de los sintagmas en la terminología presente en tesauros, con el objetivo de continuar y profundizar esta línea de trabajo. Se darán dos modalidades de análisis: una enfocada a la incidencia de determinadas estructuras morfosintácticas e idiomáticas, y otra dirigida a la extensión de los términos, medida en cantidad de palabras.

2. Metodología

Se cumplió con las siguientes fases metodológicas para desarrollar la investigación:

I. Clasificación en categorías de los sintagmas terminológicos. A partir de las apreciaciones de Sager (1993) y otros autores, se determinaron las siguientes categorías de análisis:

Categoría 1. Combinación de dos elementos lexicales en lengua española que no cuentan con ningún elemento que las enlace. En su presentación más habitual, se integran con un sustantivo y un adjetivo (National Information Standards Organization, 2005; Naumis Peña, 2007), aunque también pueden darse combinaciones del tipo “sustantivo+sustantivo”, “adjetivo+sustantivo” e incluso “adverbio+sustantivo” o “sustantivo+adverbio”. Se trata de un tipo de sintagma nominal muy productivo, del cual pueden encontrarse fácilmente ejemplos, tales como “tesauro facetado”, “coche bomba”, “primera opción” o “no directividad”.

Categoría 2. Combinación de dos sustantivos o formas sustantivadas en lengua española, conectadas por uno o más enlaces, los que suelen ser preposiciones, contracciones o preposiciones más artículos. Ejemplos típicos son expresiones como

“fondo de amortización”, “crédito al consumidor”, “sociedad de la información” o “educación de adultos”.

Categoría 3. Combinación de tres o más elementos con o sin enlaces entre sus elementos. Este procedimiento es, en realidad, un desarrollo de las construcciones clasificadas en las categorías 1 y 2. Son ejemplo los términos “síndrome respiratorio agudo severo”, “listas de encabezamientos de materia” y “control médico periódico”.

Categoría 4. Combinación de dos o más elementos lexicales, siempre que al menos uno de ellos pertenezca a una lengua diferente a la española. Esta variedad aparece regularmente en los discursos especializados, ya integrando palabras en español y lenguas predominantes como el inglés (“leasing operativo”), o ya combinando palabras en español y latín (“panel forum”). Se considerarán todos los latinismos estén o no recogidos por el DRAE. También se considerarán latinismos aquellos que se presenten en su forma latina y no en su versión española (status vs. estatus)

Categoría 5. Combinación de dos o más elementos siempre que al menos uno sea una expresión epónima o un topónimo. Conforme con Sager (1993, p. 107), los nombres propios no se comportan como nombres comunes pues no son pasibles de análisis sémico y no se emplean de forma predicativa. Muchos términos epónimos utilizan el formato de la categoría 2 (por ejemplo, “síndrome de Down” o “tetralogía de Fallot”), aunque algunos evolucionan hacia formas creadas según la categoría 1, como en “efecto Zeeman” o “doctrina Monroe”. En los lenguajes de especialidad parece darse la necesidad de combinar los valores individualizados de los nombres propios con la función clasificadora de los nombres comunes (Gutiérrez Rodilla, 1998, p. 133). En algunos casos el nombre propio se lexicaliza, como en el caso de “teoría copernicana”.

II. Selección de tesauros. Se aplicaron cuatro criterios para la selección de dos tesauros: a) elección de dos ámbitos de las Ciencias Sociales, tan distantes entre sí como fuese posible; b) calidad intrínseca de cada tesoro, evaluada a la luz de las recomendaciones de los estándares (National Information Standards Organization, 2005; British Standards Institution, 2005); c) disponibilidad en Internet; y, d) procedencia geográfica diferente. Los tesauros seleccionados en base a estos criterios son:

i) *Tesoro Europeo de Educación - TEE* (Comisión de las Comunidades Europeas, 2003). Es el resultado de más de diez años de trabajo y cooperación entre el Consejo de Europa y la Comisión de las Comunidades Europeas, y se constituye en una potente

herramienta para la indización, característica fundada en la existencia de este lenguaje en once lenguas oficiales de la Unión Europea. Cuenta con microtesauros específicos de regiones. Disponible en: <http://www.freethesaurus.info/redined/es/index.php?letra=?>

ii) *Tesaurus de Discapacidad - TD* (España. Ministerio de Sanidad Pública, Social e Igualdad, 2009). Se trata de un tesaurus estructurado y especializado en la temática de Discapacidad. Disponible en:

<http://tesaurus.observatoriodeladiscapacidad.es/index.php?letra=?>

3. Determinación de criterios y decisiones para favorecer la homogeneidad del análisis. En este marco, no se consideraron las polijerarquías presentes en el TEE. Así, cada descriptor presente en una polijerarquía se tomó en cuenta una vez, se obviaron los calificadores y descriptores repetidos, y no se consideraron los microtesauros por regiones.

En el caso del TD se excluyeron los términos cabecera (86 en total) se usan solo para agrupar descriptores y no son recomendados para la indización. También se excluyeron términos repetidos, precoordinationes, siglas, equivalencias en otras lenguas y nombres de instituciones o premios, tales como Acta Única Europea o Fondo Europeo para la Juventud. Tampoco se consideraron los calificadores del tipo (Lenguaje) en “Nivel umbral (Lenguaje)” debido a su función meramente referencial.

En cuanto a expresiones con guión del tipo “socio-cultural” fueron tomadas como una sola palabra o dos, según como fueran registradas en el DRAE. Expresiones no registradas en el DRAE (como “costo-beneficio”), se tomaron como dos palabras.

4. *Counting* y distribución de términos. El *counting* es el conteo de documentos, objetos o expresiones para obtener un dato cuantitativo (Hjørland, 2005), siempre que los elementos a ser contados sean homogéneos y pertenezcan al mismo universo estudiado. En este trabajo, los elementos contados son los términos utilizados por los tesauros, sean descriptores o no descriptores. La distribución de términos se realizó desde dos perspectivas: a) inclusión de cada uno en alguna de las cinco categorías establecidas; b) discriminación conforme a la extensión en número de palabras de cada término.

5. Establecimiento de datos cuantitativos absolutos y porcentuales.

6. Análisis de los resultados y conclusiones.

3. Resultados

El universo total de términos estudiados fue de 4.275. La tabla 1 desagrega el número de términos por tesauro y distingue entre términos simples (de una sola palabra) y compuestos (formas sintagmáticas en general), y establece los porcentajes respectivos.

Tabla 1

	TEE		TD	
	Cantidad	%	Cantidad	%
Términos simples	1.122	36.87	308	25.00
Sintagmas	1.921	63,13	924	75,00
Total	3.043	100.00	1.232	100.00

Términos discriminados cuantitativa y porcentualmente por tesauro

La tabla 2 da cuenta de la distribución de los sintagmas de los tesauros estudiados, dentro de las cinco categorías propuestas. El universo de términos estudiados suma 2.845, que corresponde al 66,55 % del total de términos (4.275) comprendidos.

Tabla 2

	TEE		TD	
	Cantidad	%	Cantidad	%
Categoría 1	1122	58.41	484	52.38
Categoría 2	639	33.26	274	29.65
Categoría 3	137	7.13	148	16.02
Categoría 4	21	1.09	2	0.22
Categoría 5	2	0.10	16	1.73
Total	1921	100	924	100

Sintagmas discriminados cuantitativa y porcentualmente por categorías

La tabla 3 también toma en consideración solo el número de sintagmas (2.845).

Tabla 3

	TEE		
--	-----	--	--

	Cant	%	Cant	%	Cant	%
Dos palabras	1140	50.34	186	52.60	213	55.07
Tres palabras	531	27.64	247	26.73	320	27.10
Cuatro palabras	207	10.78	127	13.74	167	12.26
Cinco palabras	34	1.77	35	3.70	34.5	2.78
Más de cinco	9	0.47	20	3.14	19	1.80
Total	1021	100	024	100	1422.5	100

Sintagmas discriminados por cantidad de palabras

Los resultados confirman y ajusta investigaciones anteriores respecto a dos puntos: a) la mayor parte de la terminología de un dominio se compone de sintagmas; y, b) el porcentaje de sintagmas terminológicos constituye entre un 50 y un 80 % de los términos de un área especializada. La muestra estudiada comprendió 4.275 términos de los cuales 2.845 (o sea, el 66,55 % de ellos), resultaron ser sintagmas (63,12 % de los términos del TEE y el 75 % de los del TD). Este dato es particularmente valioso si se considera la dimensión numérica de la muestra y el hecho de que la totalidad de la terminología registrada en los tesauros, ha pasado por el filtro riguroso de procesos de control de vocabulario.

La discriminación de términos por categorías permite concluir que: a) las estructuras más típicas son aquellas con combinaciones más breves; b) los términos suelen decrecer en número a medida que aumenta la cantidad de palabras que los componen; c) las dos estructuras predominantes (categorías 1 y 2) representan el 88,54 % (2.519/2845) de los sintagmas terminológicos y el 58,92 % (2.519/4.275) del total de términos estudiados; d) los términos híbridos (categorías 4 y 5) son absolutamente excepcionales, al menos en las disciplinas estudiadas; e) los datos relativos a las cinco categorías estudiadas son consistentes entre los dos tesauros. Queda por analizar la composición interna de los términos de las categorías 1 y 2, pues estudios previos permiten aventurar que las estructuras típicas suelen ser “sustantivo+adjetivo” y “sustantivo+enlace(s)+sustantivo” respectivamente.

Si se suma la cantidad de términos representados por una, dos y tres palabras, el porcentual asciende al 89,68 % (3.834/4.275). Quiere decir que solo el 10 % de los términos tiene una extensión superior a las tres palabras, lo que parece razonable y

ajustado a las recomendaciones para la creación y el desarrollo de tesauros. El análisis de términos discriminados según la cantidad de palabras que los componen, permite decir que en este caso también los datos son consistentes en los dos repertorios estudiados. A modo de ilustración, los porcentuales de términos con dos palabras se encuentran entre el 50 y el 60 % del total de sintagmas, mientras que los porcentuales de términos con tres palabras son próximos en todos los casos al 25 %. Podría considerarse la hipótesis, a confirmar en futuras investigaciones, de que el ajuste a estos porcentajes constituye uno de los indicadores de calidad de la terminología seleccionada, al menos en lo relativo a las disciplinas de las Ciencias Sociales.

4. Conclusiones

Las siguientes conclusiones constituyen una primera aproximación hacia enfoques cuantitativos para el análisis de dominios desde la perspectiva de su terminología, y hacia valoraciones cualitativas que sistematicen las modalidades de formación de sintagmas.

Cabe reconocer, en primer término, que los profesionales que trabajan con lenguaje científico y especializado, al menos en las disciplinas comprendidas en las Ciencias Sociales, deben asumir que más de la mitad de los términos se expresan a través de formas sintagmáticas, lo que tiene consecuencias metodológicas en lo que refiere a criterios de identificación, depuración y selección de términos candidatos para la indización. En el mismo sentido, los clasificacionistas deben tener presente esta realidad al momento de elaborar, revisar o evaluar vocabularios controlados.

No puede eludirse la interrogante de por qué, al menos en lengua española, siendo teóricamente posible no se suele representar conceptos en la mayoría de las áreas especializadas, con una sola palabra. No surgen de la revisión realizada en literatura bibliotecológica, investigaciones orientadas a responder claramente esa pregunta. Es posible que haya que establecer puentes con la investigación en Lingüística para llegar a conclusiones más firmes. Por otra parte, esta cuestión tiene consecuencias en el desarrollo de tesauros, ya que la alta sintagmación exige la anexión de índices permutados integrados a los esquemas de los tesauros, para permitir formas alternativas de acceso a los descriptores y no descriptores que se constituyen mediante sintagmas.

Los resultados obtenidos tienen impacto, asimismo, en el análisis de alternativas para una recuperación con bajo nivel de ruido y de fuga de información. En este punto, estos estudios se relacionan con los relativos a métodos de control de vocabulario y control de autoridades.

El análisis permite comprobar que las estructuras morfológicas de los sintagmas terminológicos se repiten a lo largo de las disciplinas y los tópicos de las Ciencias Sociales analizados en esta investigación y en otras previas, y se comportan de la misma forma en términos cuantitativos.

Se sugieren tres líneas de investigación complementaria: una orientada hacia el análisis de sintagmas presentes en otras estructuras conceptuales sistematizadoras de terminología, tales como diccionarios, glosarios y vocabularios de especialidad; otra destinada a identificar más claramente las estructuras morfosintácticas de los términos compuestos; y una tercera que desarrolle estudios similares en otras macroáreas del conocimiento (Ciencias Puras, Tecnología, Humanidades, Bellas Artes). Ello permitirá contar con datos cuantitativos pasibles de ser comparados con los presentes, y el desarrollo de nuevos elementos de evaluación.

Referencias bibliográficas

- Alves, I. M. 1998. Neologia e tecnoleitos. En Olivera, A.M.P.P. & Isquierdo, A. N, organiz. As ciências do léxico : lexicologia, lexicografia, terminologia. Campo Grande, UFMS, p. 23-29.
- Barité, M. 1997. Glosario sobre Organización y Representación del Conocimiento: Clasificación, Indización, Terminología. Montevideo: Universidad de la República, Comisión Sectorial de Investigación Científica.
- Barité, M. y Simón, L. 2010. Incidencia de sintagmas en lenguajes especializados: análisis comparativo de tesauros. En RITERM. Actas del XII Simposio Iberoamericano de Terminología. [CD-ROM]. Buenos Aires, RITERM.
- Barité, M. *et al.* 2005. Relato de un proyecto interdisciplinario de Terminología: Banco de datos terminológico sobre Contratación Comercial Internacional

Proyecto Legistern. En El Estado del arte de la Terminología y la Neología en Uruguay: ponencias y trabajos. Montevideo: Unión Latina, p. 23-36.

- Bevilacqua, Cleci. 2001. Unidades fraseológicas especializadas: novas perspectivas para sua identificação e tratamento. En Krieger, Maria da Graça y Maciel, Anna Maria Becker, orgs. Temas de terminologia. Porto Alegre/São Paulo: UFRGS ; Humanitas/USP.
- British Standards Institution. 2005. Structured vocabularies for information retrieval : BS 8723-1. S.l.: BSI.
- Cabré, M.T. 1993. La Terminología: teoría, metodología, aplicaciones. Barcelona: Atlántida; Empúries.
- Cabré, M.T.; Estopá, R. y Lorente, M. 1996. Terminología y fraseología. En V Simposio de Terminología Iberoamericana, Ciudad de México, 5-9 noviembre 1996. [Citado 12 jul 2013]. Disponible en World Wide Web: <http://www.upf.edu/pdi/df/rosa.estopa/docs/00mexic.pdf>
- Cintra, A. M. M. *et al.* 2002. Para entender as linguagens documentárias. 2a. ed. São Paulo: Editora Polis; APB.
- Comisión de las Comunidades Europeas. 2003. Tesoro europeo de la Educación. S.l.: Redined. [Citado 12 de julio de 2013]. Disponible en World Wide Web www.freethesaurus.info/redined/es/
- Corbalán Sánchez, L.M. y Amat, C.B. 2003. Vocabulario de información y documentación automatizada. Valencia: Universitat de València.
- Corpas Pastor, G. 1996. Manual de fraseología española. Madrid: Gredos.
- Dodebei, V. L. D. 2002. Tesoro: linguagem de representação da memória documentária. Niteroi: Intertexto; Rio de Janeiro: Interciência.
- Dubuc, R. 1999. Manual práctico de Terminología. 3ª ed. corregida y actualizada. Santiago de Chile: Unión Latina.
- España. Ministerio de Sanidad Pública, Social e Igualdad. 2009. Observatorio Estatal de la Discapacidad: [Tesoro]. S.l.: Junta de Extremadura. [Citado 12 de julio de 2013]. Disponible en World Wide Web <http://www.observatoriodeladiscapacidad.es/>

- Gil Leiva, I. 1999. La automatización de la indización de documentos. Gijón: Trea.
- Gil Leiva, I. y Rodríguez Muñoz, J.V. 1997. Análisis de los descriptores de diferentes áreas del conocimiento indizadas en bases de datos del CSIC: aplicación a la indización automática. Revista española de documentación científica [en línea], vol. 20, nº. 2, p. 150-160 [Citado 12 Jul 2013]. Disponible en World Wide Web: <http://webs.um.es/isgil/Descriptores%20descriptors%20indizacion%20indexing%20GIL%20LEIVA.pdf>
- Gutiérrez Rodilla, B. 1998. La ciencia empieza en la palabra: análisis e historia del lenguaje científico. Barcelona: Península.
- Hjørland, B. 2005. Core concepts in Library and Information Science (LIS). [en línea]. [Citado 12 de Jul. 2013]. Disponible en World Wide Web: <http://www.iva.dk/bh/core%20concepts%20in%20lis/home.htm>
- International Organization for Standarization. 1987. Norme internationale ISO 704 : principes et méthodes de la Terminologie. Ginebra : ISO.
- Krieger, M. DA G. y Finatto, M.J. 2004. Introdução à terminologia : teoria e prática. São Paulo: Contexto.
- Lancaster, F.W. 1996. Indización y resúmenes: teoría y práctica. Buenos Aires: EB.
- L'Homme, Marie. 2001. Combinaisons lexicales spécialisées: Regroupement des mots clés par classes conceptuelles. En Daille, B. et G. Williams, ed. Journées d'étude de l'ATALA. La collocation. Rapport de recherche. Nantes: Institut de recherche en informatique de Nantes.
- Mazzocchi, F. *et al.* 2007. Relational Semantics in Thesauri: An Overview and Some Remarks at Theoretical and Practical Levels. Knowledge Organization, vol 34, nº. 4, p. 196-213.
- National Information Standards Organization. 2005. Guidelines for the construction, format and management of monolingual controlled vocabularies: an American National Standard developed by the National Information Standards Organization. Bethesda: NISO Press. [Citado 12 Jul 2013] Disponible

en World Wide Web: http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key%3Austring

- Naumis Peña, C. 2007. Los tesauros documentales y su aplicación en la información impresa, digital y multimedia. México: UNAM.
- Real Academia Española. 2001. Diccionario de la lengua española. 22ª edición. Madrid: Espasa-Calpe. 2 v.
- Sager, J. C. *et al.* 1980. English special languages : Principles and practice in science and technology. Wiesbaden: Brandstetter.
- Sager, J. C. 1993. Curso práctico sobre el procesamiento de la terminología. Madrid: Fundación Germán Sánchez Ruipérez.